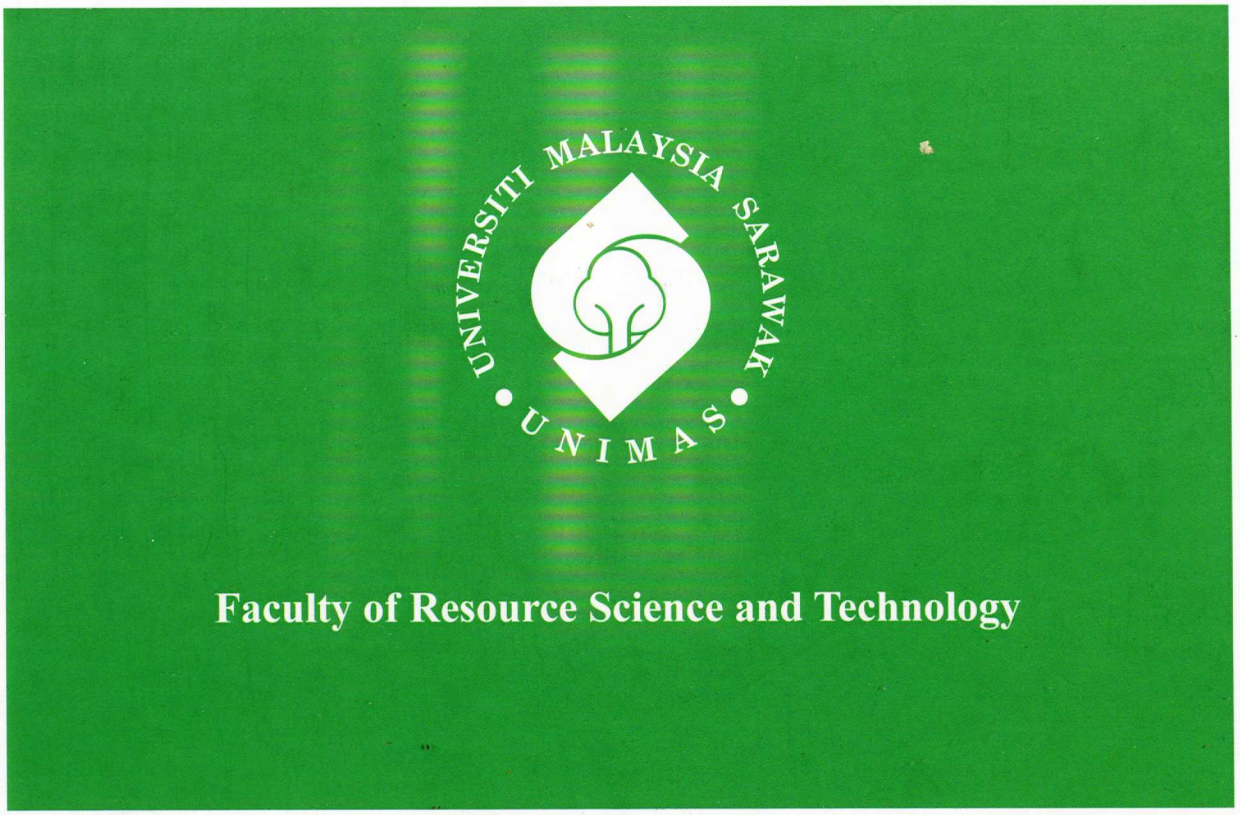

Bioinformatics Analysis of Genes involved in Trunking and Non-trunking of *Metroxylon sagu* (Sago Palm)

Philip anak Mendit
(50891)

Bachelor of Science with Honours
(Resource Biotechnology)
2018

Bioinformatics Analysis of Genes Involved in Trunking and Non-trunking of Sago Palm

Philip anak Mendit (50891)

A thesis submitted to partial fulfillment of the requirements for the Degree of Bachelor of Science with Honours

(Resource Biotechnology)

Supervisor: Associate Professor Dr. Mohd Hasnain bin Mohd Hussain

Co-supervisor: AP Dr. Hairul Azman

Department of Molecular Biology

Faculty of Resource Science and Technology

Universiti Malaysia Sarawak

2018

UNIVERSITI MALAYSIA SARAWAK

Grade: _____

Please tick (√)
Final Year Project Report ☒
Masters ☐
PhD ☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the 12 day of June year 2018

Student's Declaration:

I PHILIP ANAK MENDIT 50891 FACULTY OF RESOURCE SCIENCE AND BIOTECHNOLOGY

(PLEASE INDICATE NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled, _____ is my original work. I have not copied from any other students' work or from any other sources with the exception where due reference or acknowledgement is made explicitly in the text, nor has any part of the work been written for me by another person.

12 JUNE 2018
Date submitted

PHILIP ANAK MENDIT (50891)
Name of the student (Matric No.)

Supervisor's Declaration:

I,---------- (SUPERVISOR'S NAME), hereby certify that the work entitled, ----------(TITLE) was prepared by the aforementioned or above mentioned student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of ---------- (PLEASE INDICATE THE DEGREE TITLE), and the aforementioned work, to the best of my knowledge, is the said student's work

Received for examination by: _____ Date: _____
(Name of the supervisor)

I declare this Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*
☑ **OPEN ACCESS**

I declare this Project/Thesis is to be submitted to the Centre for Academic Information Services (CAIS) and uploaded into UNIMAS Institutional Repository (UNIMAS IR) (Please tick (√)):

☑ **YES**
☐ **NO**

Validation of Project/Thesis

I hereby duly affirmed with free consent and willingness declared that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abide interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic and research purposes only and not for other purposes.
- The Centre for Academic Information Services has the lawful right to digitize the content to be uploaded into Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis if required for use by other parties for academic purposes or by other Higher Learning Institutes.
- No dispute or any claim shall arise from the student himself / herself neither a third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student himself/herself without first obtaining approval from UNIMAS.

Student's signature _____ 12/06/2018 (Date) Supervisor's signature: _____ (Date)

Current Address:
PO BOX 54 NO 1/97, KM 2 1/2, SIMPANG PAIP, JALAN SARATOK, 95400 SARATOK

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the date of restriction indicated, and the reasons for the confidentiality and restriction.

[The instrument was prepared by The Centre for Academic Information Services]

ACKNOWLEDGEMENT

My heartfelt gratitude towards my supervisor, Associate Professor Dr. Mohd Hasnain bin Mohd Hussain that had guided me throughout this study. Thank you for all the ideas and aids in terms of sharing of knowledge and providing me with the information that helped me a lot during this research.

Thanks to Yan Wei Jie, the postgraduate student for his continuous help and guidance throughout this whole research. Thank you for introducing me with the wonderful world of bioinformatics and sharing of knowledge in this field.

I would also like to thank my mother and father, Mr. Mendit and Mrs. Catherine for their support, especially by providing me with financial support throughout this whole project. Thank you also to my best friend, Darren Richmond for providing me with financial support in case of emergency while I am completing my final year project.

Thank you also to those that helped me throughout my study and upon completing this research.

Bioinformatics Analysis of Genes involved in Trunking and Non-trunking of Sago Palm

Philip anak Mendit

Resource Biotechnology
Faculty of Resource Science and Technology
University Malaysia Sarawak

ABSTRACT

Sago palm or its scientific name, *Metroxylon sagu* had been identified as the source of carbohydrates in Southeast Asia, especially in rural areas. Sago palm is being exploited mainly for their starch and was reported as the first palm used by man in South-East Asia and Oceania. This research aims to understand how non-trunking of *M. sagu* occurred at molecular level. From the findings of this research, factors affecting the growth and development of a plant can be studied to improve the quality of crops in the future. This research study the methods to compare the previously identified genes of trunking and non-trunking of *M. sagu*. Bioinformatics online databases were used in which the sequences were identified using pBLAST. Furthermore, pathways of the enzymes involved the previously identified genes were studied using KEGG. From the analysis using bioinformatics analysis, metabolic pathways were found out to be the pathways that governing the growth and development of *M sagu*, thus affecting the formation of its trunk.

Keywords: sago palm, non-trunking, bioinformatics, proteins, pathways, growth, development

ABSTRAK

Pokok sagu ataupun nama saintifiknya, Metroxylon sagu *telah dikenal pasti sebagai sumber karbohidrat di Asia Tenggara, terutamanya di kawasan pedalaman. Pokok sagu telah dieksploitasi untuk kanjinya dan dilaporkan sebagai palma pertama yang digunakan oleh manusia di Asia Tenggara dan Oceania. Kajian ini mensasarkan untuk memahami kejadian tidak berbatang* M. sagu *pada tahap molekul. Hasil daripada kajian ini, factor yang mempengaruhi pertumbuhan dan perkembangan pokok sagu dapat dipelajari untuk meningkatkan kualiti tanaman pada masa akan datang. Kajian ini mempelajari cara-cara untuk membezakan susunan gen* M. sagu *yang berbatang dan tidak berbatang yang telah dikenal pasti sebelum ini. Pangkalan data bioinformatik atas talian telah digunakan yang mana pBLAST digunakan untuk mencari persamaan pada susunan gen pokok sagu yang berbatang dan tidak berbatang. Tambahan lagi, laluan enzim yang terlibat dengan susunan gen* M. sagu *yang berbatang dan tidak berbatang yang telah dikenal pasti sebelum ini telah dikaji menggunakan KEGG. Hasil daripada analisis menggunakan alat bioinformatik, laluan metabolic telah dijumpai sebagai laluan yang menerajui pertumbuhan dan perkembangan* M. sagu, *sekali gus mempengaruhi pembentukan batang pokok sagu.*

***Kata kunci**: pokok sagu, tidak berbatang, bioinformatik, protein, laluan, pertumbuhan, perkembangan*

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PCR	Polymerase Chain Reaction
Al	Aluminium
Fe	Ferum
DNA	Deoxyribonucleic acid
Mn	Manganese
RAPD	Random Amplified Polymorphic DNA
CTAB	Cetyl trimethylammonium bromide
SNARE	Soluble NSF attachment protein receptors
EST	Expressed sequence tags
RT-PCR	Real time-polymerase chain reaction
NCBI	National Centre of Biotechnology Information
cDNA	Complementary deoxyribonucleic acid
RDA	Representational Difference Analysis
KEGG	Kyoto Encoyclopedia of Genes and Genomes
pBLAST	protein Basic Local Alignment Sequence Tool
mRNA	messenger riboxynucleic acid

INTRODUCTION

Sago palm or its scientific name, *Metroxylon sagu* had been identified as the source of carbohydrates in Southeast Asia, especially in rural areas (Abbas *et al.*, 2017). Sago palm is being exploited mainly for their starch and was reported as the first palm used by man in South-East Asia and Oceania (Ave, 1977). Sago palm produces a huge amount of starch and it is estimated that between 100 and 400 kilograms of starch are obtained from a single palm. Sago palm stores its starch in the pith in the trunk of the sago palm.

Research on *M. sagu* is vital as apart from overcoming the problems of depletion of renewable resources worldwide, *M. sagu* could be the answer towards overcoming the issues of food security. This study focuses on investigating the reasons behind the occurrences of non-trunking *M. sagu* and comparing it with the normal, trunking *M. sagu* at molecular level. Thus, this study focuses on what affects the formation of trunk of *M. sagu* and what causes them to form the normal and trunking, and the non-trunking *M. sagu*. This can be an important foundation to understand more on factors affecting the growth and development of plants. The non-trunking of *M. sagu* is a major concern in sago plantation as this phenotype forms of *M. sagu* eliminates the value of the plant economically. Since that *M. sagu* could become an answer to continuous supply of food in the future, this study is another platform to improve the starch production of *M. sagu* even in the non-trunking *M. sagu*. This research and some previous researches on trunking and non-trunking of *M. sagu* are important outbreak as *M. sagu* can be used as emerging major food source to reduce the number of global population suffering from insufficient food supply and malnourish. Thus, the study on *M. sagu* at molecular level is an effort to enhance agricultural productivity.

Generally, bioinformatics analysis is a field in biological and computing science in which it involves the science of information. In this study, bioinformatics is used to analyse the sequences of trunking and non-trunking obtained from previous works. Bioinformatics in other words is a tool for analysis of molecular biology to provide information on studying plants at their genomics and proteomics level. In this study itself, pBLAST and KEGG are the main platforms for analysing the sequences of trunking and non-trunking of *M. sagu*. The results from pBLAST and KEGG were interpreted to achieve the goal of understanding the occurrences of trunking and non-trunking of *M. sagu*.

Previous works had been done and those works were utilized in terms of foundational ideas on how this study could lead to a better understanding on development and growth of *M. sagu*. One of the key ideas or research questions in this study is how to compare the previously identified partial genes of trunking and non-trunking of *M. sagu* with the sequences of genes of bioinformatics online databases. *M. sagu* is a not a model plant and genome sequencing of this plant had not yet been done. Thus, this study or research discusses and presents the method or alternative that can be applied to elucidating the previously identified partial genes of trunking and non-trunking of *M. sagu*. Another key idea governing this study is what are the pathways of the enzymes involved the genes of trunking and non-trunking of *M. sagu*. The genes that code the proteins, for the example the enzymes, in the *M. sagu* are expressed in the pathways that related to the growth and development of *M. sagu*. Early in the study, it was hypothesized that both the genes involved in the trunking and non-trunking of *M. sagu* has different pathways and analysis with bioinformatics tools will relate these genes with their corresponding pathways. With regards to the hypothesis, the study aimed to determine the pathways of the identified trunking and non-trunking genes of *M. sagu*.

At the end of the study, most of the genes or previously identified sequences of trunking and non-trunking of *M. sagu* code for most of the proteins existed in the online databases. And some of these proteins were found to be involved in certain pathways that both related and not related to the growth and development of *M. sagu*. Further analysis found out that metabolic pathways were overlapped between the previous two works on trunking and non-trunking of *M. sagu*.

LITERATURE REVIEW

Sago Palm

Sago palm or its scientific name, *Metroxylon sagu* had been identified as the source of carbohydrates in Southeast Asia, especially in rural areas (Abbas *et al.*, 2017). It belongs to the order *Arecales Nakai*, family *Palmae Jussieu* and genus *Metroxylon Rottboell* (Dransfield *et al.*, 2005). Sago palm is being exploited mainly for their starch and was reported as the first palm used by man in South-East Asia and Oceania (Ave, 1977). Sago palm produces a huge amount of starch and it is estimated that between 100 and 400 kilograms of starch are obtained from a single palm. Sago palm stores its starch in the pith in the trunk of the sago palm. Basically, the starch from its trunk is obtained by cutting out the pith from the trunk and kneaded and washed to get the starch.

Figure 1. The *Metroxylon sagu* in its natural habit. Retrieved from http://www.palmpedia.net/wiki/Metroxylon_sagu

The study on sago palm is important as sago palm can be the most formidable source of starch for the future. As being reported by Evra Raunie in 2013, below are the potential futures for sago palm based on the studies from several reports:

1. Sago palm is harvestable throughout the year because it has suckers as their successor palms and therefore requires no replanting.

2. Sago palm is a robust crop as it prefers moist soil and has moderate tolerance to salinity. Other than that, sago palm also tolerates short period of drought, flooding and able to grow in marginal soil. Low pH, high Al, Fe and Mn in the soil are toxic to most of the plants but sago palms able to tolerate these factors.

3. Sago is focussed as food sources and industrial starch-based industries as it can be manufactured to produce food, pharmaceutical products, and the modification of starch for industrial purposes.

4. Sago waste is cellulosic biomass, woody components and water (Evra Raunie, 2013). It can be utilized, processed and turned into useful end products such as biodegradable plastics, bio-fuels, ruminant feed fibre boards.

Trunking and Non-trunking of *M. sagu*

A trunking sago palm is a normal sago palm. The trunk is full of starch at its maturity stage. Scientifically, the depth of the peat affects the formation of trunk as the normal or good conditions of trunking palms are on shallow peat (Hussain *et al.*, 2012b). Geographically, the soil of trunking sago palms area is more humidified as compared to those in the non-trunking sago palms area (Karim *et al.*, 2008). The trunking sago palm also

requires extra-moistened soil as this soil, around the trunking palms with elevated mass compactness enhances nutrient assimilation by roots. This is to improve the soil-root bond.

On the other hand, a non-trunking sago palm is sago palm with abnormal growth and phenotypic appearances with stunted growth and development. This non-trunking *M. sagu* did not form trunk even after been through 8-year cycles (Shera & Edward, 2013), thus devoid of starch in its trunk. The non-trunking of *M. sagu* is a major concern in sago plantation as this phenotype forms of *M. sagu* eliminates the value of the plant economically. Geographically, non-trunking sago palms are all grown on deep peat. The soils of non-trunking sago palms are less humidified, leading to poor growth and development, very high mortality rate and poor ability to develop trunk (Shera & Edward, 2013).

However, researches on sago palm are still in its infancy, even though there are efforts to study *M. sagu* on molecular level. Apart from that, some studies on *M. sagu* also involved the investigation of the factors that affect the accumulation of starch. By relating the factors that affect the accumulation of starch from previous studies with factors affecting plant growth and development, it is possible to track back the ideas behind the occurrences of trunking and non-trunking *M. sagu.* The study and research on trunking and non-trunking of *M. sagu* is important because it may contribute to the understanding on how the expression or suppression of certain genes affects the production of proteins that involve in plant growth and development.

The expression of genes in the trunking and non-trunking of *M. sagu* had been identified as the key or the main idea to overcome the problems in terms of growth of development in plants. This is supported by the fact that the development and the growth of *M. sagu* are regulated by the expression of the genes involved in metabolic pathways. Hence, by comparing the genes of trunking and non-trunking of *M. sagu* with the sequences of the existed genes in bioinformatics online databases, the difference in the gene expression can

be investigated and interpreted.

Physiology of *M. sagu*

Even though there are countless of researches involving the physiology of plants kingdom, there are limited information on the physiological study of *M. sagu*. In theory, the study on the growth response of *M. sagu* to environmental stresses can develop a foundational or building ideas towards investigating the factors affecting the physiology of *M. sagu* (Kamal, 2014). In fact, the difficulties to get uniform plant materials due to low germination percentage of *M. sagu* seed and large variation in days for germination remain a major gap in stress tolerance in *M. sagu*, thus explained the limited number of existing studies (Ehara, 2009).

In previous studies, *M. sagu* is considered to be salt-tolerance. Recent study reported that the ability of *M. sagu* to tolerate salt might consist of salt avoidance by mechanical restriction of excessive sodium ions distribution to the cortex (Ehara *et al.,* 2009). Apart from that, another important aspect of *M. sagu* physiological property is the photosynthetic ability of their leaves. This explained their well adaptation to the soils with poor nutrients (Matsumoto *et al.,* 1998).

Bioinformatics Analysis of Plant Genomes

Generally, bioinformatics analysis is a field in biological and computing science in which it involves the science of information. The information scientifically flows in biological systems and applying the computational methods in genetics and genomics. For example, the knowledge in bioinformatics is currently used in the analysis of microarray

data to specifically measure the alternative splicing genome-wide (Lesk, 2002). Bioinformatics analysis is a crucial technique in genetics and genomes analysis. Currently, bioinformatics analysis is a new tool or measurement technique that able to produce large quantities of biological data. In this technology era, it is significant to talk about biological information without considering the emergence of computing science in biology. Hence, it is a paradigm shift in biology to compromise bioinformatics in biological analysis.

Bioinformatics analysis covers a very wide field in life sciences and technologies. It is an interdisciplinary field involving both molecular biology and genetics, aided by computer sciences, mathematics and statistics. The analysis in bioinformatics works based on sequences analysis, structure analysis and microarray data analysis. Previously, bioinformatics contributed a lot in the breakthrough in molecular biology. For example, the work on the comparative analysis of plant genomes (Vedovato *et al.*, 2009) enabled the definition of Phytolongins, a novel non-SNARE longin domain protein family. Some other previous work involved bioinformatics was the elucidation of gene structure and function of prokaryotes and eukaryotes via DNA information technology (Hansen *et al.*, 2014).

This research utilizes bioinformatics to analyse the differences at both proteomics and genomics level of trunking and non-trunking *M. sagu.* Previous studies on *M. sagu* applied the wet laboratory technique in which experimental techniques involving the isolated genes of trunking and non-trunking of *M. sagu* such as PCR of those genes, RAPD of the genes and some others (Kamal, 2014). By studying *M. sagu* via bioinformatics analysis, this can be an elementary step of understanding on what contributes to trunking and non-trunking of *M. sagu.* For instance, microarray technology has now becoming a common bioinformatics tool in investigating gene expression. Microarray in fact is a powerful bioinformatics tool involved in the analysis of plant genomics (Skuse & Du, 2008). In this research itself, pBLAST is utilized to find the similarity between the previously identified

sequences of trunking and non-trunking genes of *M. sagu* with those existed proteins from the online databases. By employing this critical step in this research, the sequence similarity across genomes can be identified, the right primers for future study of trunking and non-trunking of *M. sagu* can be determined and any mutations in the previously identified sequences can be detected.

Previously, bioinformatics analysis was utilized in some researches involving sago palms. RAPD markers were used to natural pollinate the *M. sagu* to reveal the genetic variation of the *M. sagu* progenies (Abbas *et al.*, 2017). Bioinformatics also had been used as a molecular biology tools that acts as the platform to provide more information on the *M. sagu.* Apart from that, another study also had been conducted regarding the construction of the expressed sequence tags of *M. sagu.* The establishment of EST databases was an approach to accelerate the studies and researches of non-model and emerging species such as the *M. sagu* (Ching *et al.*, 2012).

In the research of *M. sagu*, bioinformatics tools are utilized to predict the functions of genes of the trunking and non-trunking of *M. sagu.* This can be achieved by studying and investigating the significant differences and similarities among species. Apart from that, utilizing bioinformatics in the research *M. sagu* by comparing it within plant species will allow the association of the genes function (Ong *et al.*, 2016). In this research, bioinformatics tools, with the inclusion of the online databases of the bioinformatics, are utilized for the comparative analysis of the trunking and non-trunking genes of *M. sagu*. This comparative analysis also applying the concepts of co-expression analysis (Usadel *et al., 2009).*

Previous Works on Trunking and Non-trunking of Sago Palm as basis to this Study

The previous works on trunking and non-trunking of sago palm are described below. These two works produced important breakthroughs as the sequences of amino acids used by this study are obtained from these two works. This research and some previous researches on trunking and non-trunking of *M. sagu* are important outbreak as *M. sagu* can be used as emerging major food source to reduce the number of global population suffering from insufficient food supply and malnourish. Thus, the study on *M. sagu* at molecular level is an effort to enhance agricultural productivity.

Representational Difference Analysis(RDA)

RDA is a very influential approach for identifying DNA sequence differences that associated with changes in phenotype and disease (Lisitsyn, 1995). In other words, RDA involves in the positional cloning of genes of interest (Lisitsyn, 1995). Subtractive hybridization was the first general method involved in the identification of sequences present in only one of two highly related DNA samples. In general, a typical subtraction experiment involves the mixing of two DNA samples that had been fragmented, denaturation by heat or alkali to form separate strands and reformation of hybrid double helices by reannealing the complementary strands (Lisitsyn, 1995). It is important to understand that in most of the applications of RDA, one of the DNA samples known as tester is mixed with a large excess of another DNA sample known as driver (Lisitsyn, 1995). This is vital as to corroborate that the tester DNA fragments hybridized predominantly with the driver DNA fragments. In certain experiments, if the driver DNA fragments are labelled with a hapten, the driver DNA and the hybrid DNA can be subtracted by affinity chromatography and resulting in the

purification of the target DNA fragments (Lisitsyn, 1995). This purified target DNA fragments are present only in the tester DNA sample. In advance, this method was experimented successfully to clone cDNA copies of transcripts that are differentially expressed in one of two cell types.

RDA is being discussed as the amino acid sequences of trunking and non-trunking of *M. sagu* were obtained from the study. In RDA for Identification of Molecular Factors Contributing to Trunking and Non-trunking *M. sagu*, the study aimed to identify the differences between the trunking and non-trunking *M. sagu* transcriptome (Kamarol, 2015). Subtractive and kinetic enrichment were applied in the technique to separate the difference fragment between two populations of DNA. Then, in this study, the two DNA genomes or cDNA populations were compared and then hybridized, to remove the common sequence of the genomes or cDNA populations (Kamarol, 2015). Amino acids sequences obtained from this study were the core components for illustrating the pathways of those sequences in trunking and non-trunking *M. sagu*.

In general, the methods involved in RDA are representation and difference enrichment. RDA efficiently combines subtractive hybridization with another two additional elements: representation and kinetic enrichment (Lisitsyn, 1995). By definition, representation in RDA refers to the highly reproducible sampling of two genomic DNAs, involving digestion on the DNAs with a restriction endonuclease to generate fragments with an average length of 2-5 kb, followed by ligation to oligonucleotide adaptors and whole genome PCR of short fragments only (Lisitsyn, 1995). However, representation brings out two important effects. First, representation results in the driver and tester DNA samples having reduced sequence complexity. This is because most of the DNA fragments are too long to be amplified. The efficiency of subsequent subtractive hybridizations is increased because of this. Second, representation converts differences in the restriction sites between

the two genomic DNAs (restriction fragment length polymorphisms) to differences in the DNA sequence content between the driver and the tester DNA samples. Hence, binary polymorphisms can be detected by probes and can be readily isolated.

On the other hand, difference enrichment combines subtractive hybridization with kinetic enrichment and provides an enormous purification of the target DNA sequence (Lisitsyn, 1995). All the tester DNA fragments are ligated to non-phosphorylated oligonucleotide adaptors before each round of purification of the target DNA sequence so that they have a long oligonucleotide at their 5' ends. A short oligonucleotide is used to raise the efficiency of ligation and is not covalently attached and dissociates from the DNA at elevated temperatures. Each round starts by mixing the tester DNA sample and then denaturation and reassociation of the mixture. The oligonucleotide cohesive ends are filled with Taq DNA polymerase, followed be PCR using the long oligonucleotide as the primer. An important point to ponder is the target DNA fragments are absent in the driver DNA sample. These target DNA fragments are predominantly self-reannealing and later form homoduplexes with oligonucleotide sequences on both ends and thus, the target DNA fragments are selectively amplified. Simultaneously, the single strands of non-target tester fragments predominantly hybridize or form heteroduplexes with driver DNA fragments (Lisitsyn, 1995). The tester-driver hybrid DNA duplexes fail to participate in the exponential amplification as they have the primer sequence on one end only and are therefore subtracted.

In scientific researches, RDA are applied in detection of genetic lesions in cancer, discovery of new pathogens and isolation of polymorphic markers linked to a trait. Apart from that, RDA had been successfully used to generate markers for differentiation between varieties of date palms (*Phoenix dactylifera*), which is vital as the results from the research can be analysed and relate it with this current study. In advance, this method is different from other methods that have been used for identification of varieties as some other

researches relied on comparison of presence and absence of band such as RAPD (Kamarol, 2015). In plant genomic analysis, RDA allows the take out of the right sequences that differ between two genomes from closely related plant.

Development of RDA was carried out by using cDNA as a representation instead of whole DNA genome (Kamarol, 2015). In this analysis, mRNA is used as starting material to generate cDNA followed by ligation of adaptor and PCR amplification. The main advantage of using cDNA RDA method is that it can identify the difference in gene expression between two populations and over expressed genes in one population can be isolated followed by cloning. Moreover, differently expressed genes from samples can be identified based on its growth stages. At the same time, this method can detect the up and downregulated genes.

In the study from which the amino acid sequences of trunking and non-trunking *M. sagu* were obtained for this bioinformatics analysis study of those sequences, cDNAs of trunking and non-trunking of *M. sagu* were compared to identify the difference in gene expression (Kamarol, 2015). These cDNAs were hybridized and amplified to isolate the differences in the transcriptome and or expression of genes. cDNAs of non-trunking *M. sagu* were designated as tester DNA while the cDNAs of trunking *M. sagu* were designated as driver to identify the overexpression of the genes of non-trunking *M. sagu*. In the study, tester cDNA contained sequence of interest and driver cDNA was a normal sample. A set of adaptors was used to ligate the tester DNA and was hybridized with excess of driver followed by amplification. Elimination of common sequences was don e and the sequence of interest was amplified exponentially.

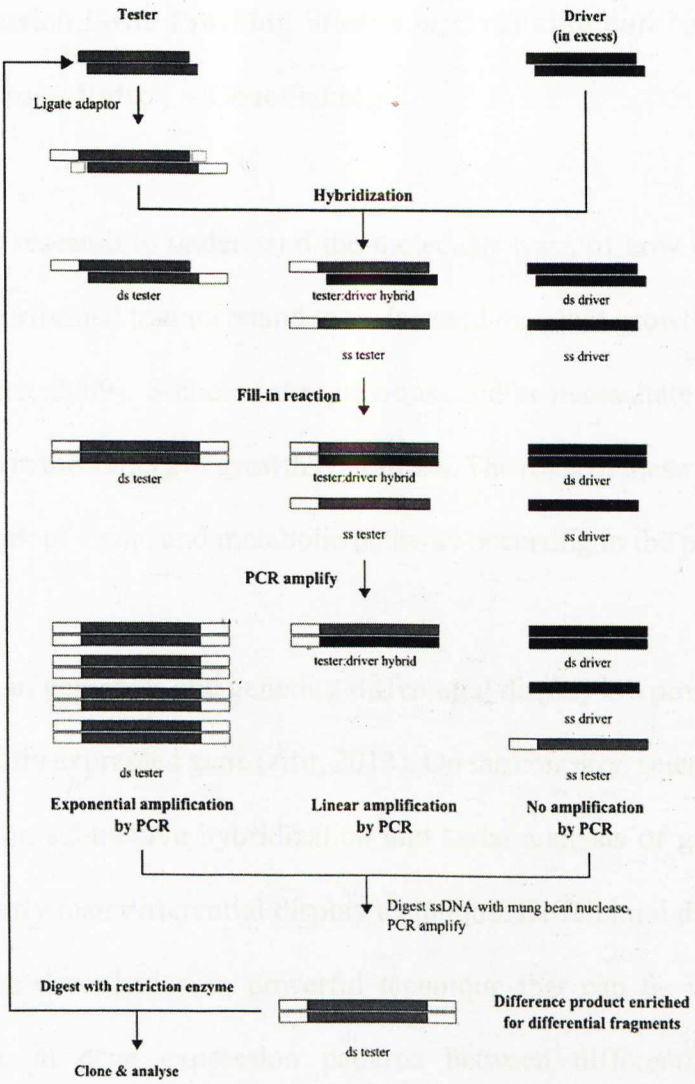

Figure 2. The general overview on how Representational Difference Analysis technique is being carried out. Retrieved from https://www.researchgate.net/figure/Fig-2-Representational-difference-analysis-Double-stranded-DNA-tester-driver-is_fig2_223516009